

这里我们找出法方程的来源。设 $f(X) = f(x_1, \dots, x_n)$ 是一个多元函数，回顾多元微积分里求极值的标准步骤：

(1) 求驻点，即解方程：

$$\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2} = \dots = \frac{\partial f}{\partial x_n} = 0.$$

(2) 对每一个驻点 $M = (a_1, \dots, a_n)$ ，判断Hessian矩阵 $H(f)|_M$ 的正定性。

下面我们说明最小二乘问题里提到的法方程来自于上面的步骤一。对于线性方程组 $AX = \beta$ ，我们想找 X ，使得 $(AX - \beta)^T(AX - \beta)$ 最小。令

$$f(X) = f(x_1, \dots, x_n) = (AX - \beta)^T(AX - \beta).$$

求最小二乘解的问题即是求 $f(X)$ 的极小值的问题。显然最小二乘解是 f 的驻点，我们的问题是：怎么样求函数 f 的驻点？

我们把 f 的表达式变换一下形式

$$\begin{aligned} f(X) &= (AX - \beta)^T(AX - \beta) = (X^T A^T - \beta^T)(AX - \beta) \\ &= X^T A^T A X - \beta^T A X - X^T A^T \beta + \beta^T \beta. \end{aligned}$$

把 AX 看成一个整体，这是关于 X 的一个向量函数：

$$AX = \begin{pmatrix} f_1(x_1, \dots, x_n) \\ \dots \\ f_n(x_1, \dots, x_n) \end{pmatrix} = \begin{pmatrix} a_{11}x_1 + \dots + a_{1n}x_n \\ \dots \\ a_{n1}x_1 + \dots + a_{nn}x_n \end{pmatrix}.$$

直接计算，或者利用公式https://en.wikipedia.org/wiki/Matrix_calculus#Scalar-by-vector_identities，我们有

$$\begin{aligned} \frac{\partial f(X)}{\partial x_i} &= \frac{\partial X^T A^T A X}{\partial x_i} - \frac{\partial \beta^T A X}{\partial x_i} - \frac{\partial X^T A^T \beta}{\partial x_i} + \frac{\partial \beta^T \beta}{\partial x_i} \\ &= 2X^T A^T \begin{pmatrix} a_{1i} \\ \dots \\ a_{ni} \end{pmatrix} - 2\beta^T \begin{pmatrix} a_{1i} \\ \dots \\ a_{ni} \end{pmatrix}. \end{aligned}$$

将驻点条件带入上面的方程，得到对任意 i ，

$$2X^T A^T \begin{pmatrix} a_{1i} \\ \dots \\ a_{ni} \end{pmatrix} - 2\beta^T \begin{pmatrix} a_{1i} \\ \dots \\ a_{ni} \end{pmatrix} = 0.$$

然后将 n 个方程拼在一起，得到

$$2X^T A^T A - 2\beta^T A = \theta^T.$$

两边取转置即得到课本上定义的法方程。